Exploiting Diffusion Prior for Real-World Image Super-Resolution

Abstract

We present a novel approach to leverage prior knowledge encapsulated in pre-trained text-to-image diffusion models for blind super-resolution. Specifically, by employing our time-aware encoder, we can achieve promising restoration results without altering the pre-trained synthesis model, thereby preserving the generative prior and minimizing training cost. To remedy the loss of fidelity caused by the inherent stochasticity of diffusion models, we employ a controllable feature wrapping module that allows users to balance quality and fidelity by simply adjusting a scalar value during the inference process. Moreover, we develop a progressive aggregation sampling strategy to overcome the fixed-size constraints of pre-trained diffusion models, enabling adaptation to resolutions of any size. A comprehensive evaluation of our method using both synthetic and real-world benchmarks demonstrates its superiority over current state-of-the-art approaches. Code and models are available at https://github.com/IceClear/StableSR.